energies

MDPI

Article

# Optimal Design of Wireless Charging Electric Bus System Based on Reinforcement Learning

**Hyukjoon Lee, Dongjin Ji and Dong-Ho Cho ***

KAIST (Korea Advanced Institute of Science and Technology), 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Korea; reloadingmemory@kaist.ac.kr (H.L.); jdj0524@kaist.ac.kr (D.J.)

* Correspondence: dhcho@kaist.ac.kr

Received: 16 February 2019; Accepted: 27 March 2019; Published: 30 March 2019


**Abstract:** The design of conventional electric vehicles (EVs) is affected by numerous limitations, such as a short travel distance and long charging time. As one of the first wireless charging systems, the Online Electric Vehicle (OLEV) was developed to overcome the limitations of the current generation of EVs. Using wireless charging, an electric vehicle can be charged by power cables embedded in the road. In this paper, a model and algorithm for the optimal design of a wireless charging electric bus system is proposed. The model is built using a Markov decision process and is used to verify the optimal number of power cables, as well as optimal pickup capacity and battery capacity. Using reinforcement learning, the optimization problem of a wireless charging electric bus system in a diverse traffic environment is then solved. The numerical results show that the proposed algorithm maximizes average reward and minimizes total cost. We show the effectiveness of the proposed algorithm compared with obtaining the exact solution via mixed integer programming (MIP).

**Keywords:** reinforcement learning; wireless charging electric bus system; Markov decision model; optimization; Q-learning

## 1. Introduction

Electric vehicles are soon likely to replace those powered by an internal combustion engine due to their high efficiency and low pollution. This research trend has encouraged many governments around the world to announce new policies aimed at replacing automobiles powered by internal combustion engines with electric vehicles. However, there are still a number of critical drawbacks in the use of electric vehicles for which solutions must be found. First, the battery in an electric vehicle occupies more space than it does in a gasoline or diesel engine, and its limited capacity implies shorter travel distances between recharging points. Moreover, battery charging time is longer than refuelling time for a conventional vehicle. This is a major problem for drivers who must therefore spend more time at charging stations.

The Online Electric Vehicle (OLEV) was introduced by KAIST [1] in an attempt to overcome these problems. In this new wireless charging system, power is drawn from underground power cables, meaning that a large battery is not necessary because the motor instantaneously receives power from the embedded power cables, which also leads to improved energy efficiency in the vehicle due to its lighter weight. Furthermore, drivers do not need to spend time charging their vehicles because the installed power cables continuously transmit energy, and the vehicle is thus charged while it is being driven along. For all these reasons, we believe that wireless charging will characterize the next generation of transportation systems, eventually replacing conventional electric vehicles altogether. As shown in Figure 1, the OLEV itself appears very similar to a conventional electric vehicle, but significantly differs because a power-receiving pickup module is part of the on-board

equipment contained in the vehicle, in addition to a motor and a battery. The power-receiving unit is attached to the bottom of the vehicle and picks up the transmitted power from the power cable; the regulator then supplies a constant voltage to the battery. The power cable is installed beneath the surface of the road on which the vehicle operates. The power-supply infrastructure is composed of an inverter and a power cable installed in segments.

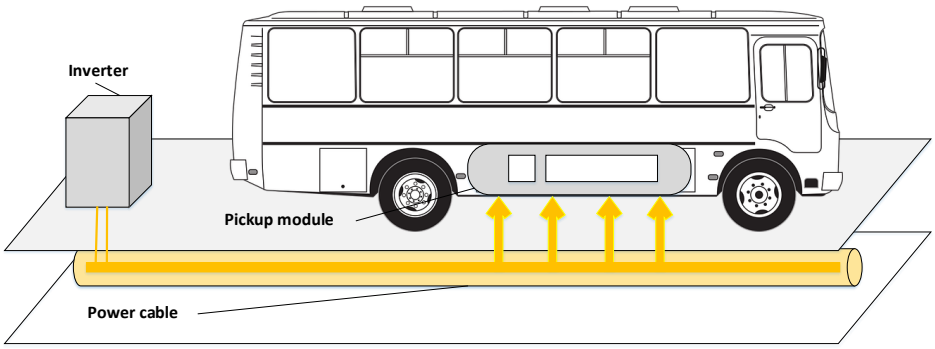


**Figure 1.** Overall layout of wireless charging system.

The technical issues associated with the wireless charging of an electric vehicle have previously been described. These include double-sided inductor–capacitor–capacitor compensation, electric-field resonance, and magnetic shielding, all of which relate to the efficiency and safety of the vehicle's wireless charging [2–4]. A study [5] investigated the various designs of transmitting and receiving coils for on-road charging systems. Different circuit models were introduced based on the number of charging EVs, and a power-control algorithm was presented to respond to diverse traffic environments. Kim et al. [6] proposed a high-efficiency coil-design formula applied to a wireless power-transfer system for a 1 kW golf cart. The system achieved 96% power-transfer efficiency for an operating frequency of 20.15 kHz, and a 156 mm air gap between the coils. The proposed coil design makes high power transfer possible for commercial electric vehicles. A commercial version of the OLEV was developed by KAIST [7], and is now operating successfully in Kumi, South Korea. Other authors considered the question of optimization related to power-transmitter positioning and battery-capacity determination; in their research, these authors proposed a logical approach to the trade-off between battery size and the positions of the power transmitters along the route of the vehicle [8–10]. The cost of the logistics for three different types of charging systems, stationary, quasidynamic, and dynamic, was analyzed, and cost-sensitivity analysis was performed [11]. The number of power transmitters and the location of the installations were selected as decision variables, and were then optimized based on an mixed integer programming (MIP)-based heuristic algorithm [12–14]. An optimization model was used to minimize the total cost of installing the power transmitters. These authors developed an MIP model to locate a charger for plug-in electric vehicles [15,16]. In order to properly allocate charging stations for Plug-in Electric Vehicles, the MIP-based model was used to evaluate the trip-success ratio to increase charging-station accessibility for drivers [17]. The MIP-based model is structured in two stages: first, estimating the range between the driver and the available charging station and, second, allocating the charging station. The study was then expanded to consider the allocation of wireless power-transfer chargers during operation [18], using dynamic programming to achieve optimal allocation of the wireless power-transfer system. Reference [19] demonstrates the effects of on-road charging on the range of EV travel for varying levels of power transmission. The simulation was conducted using a standard driving cycle, which was designed to reflect urban and highway driving scenarios (i.e., UDDS, HWFET2).

Conventional studies regarding the optimization of wireless charging systems have several limits. First, most studies are simulated under a static traffic environment, which evaluates the EV's travel distance based on the designated driving cycle. Then, there is a limitation in reflecting a real traffic environment, because traffic flow changes over time in actual traffic environments. To precisely

evaluate EV performance, a simulation under dynamic traffic environments, built based on real-time traffic data, is crucial. In the case of a dynamic traffic environment, the MIP-based exact algorithm's computational complexity greatly increases as the number of constraints escalates [20]. Moreover, the MIP-based exact algorithm needs to be modified according to traffic environment changes, which makes it extremely inefficient to find the optimal solution in dynamic traffic environments.

In order to efficiently optimize the wireless charging electric bus system in a dynamic traffic environment, the reinforcement-learning algorithm approach is an attractive alternative compared to the MIP-based exact solution. When the MIP-based exact solution needs to be modified after receiving new traffic information, the proposed algorithm can adapt to diverse traffic environments. This is possible because it does not require any prior knowledge of traffic changes in the environment, which makes it suitable for a system based on real-time data without any future information. In the present paper, our main aim is to optimize the main parameters of a wireless charging electric bus system using a proposed wireless charging electric bus system model and an optimization algorithm based on reinforcement learning. The contributions of this work are as follows:

- We propose a precise model of a wireless charging electric bus system based on a Markov decision process (MDP), which is composed of environment, state, action, reward, and policy.
- For accurate analysis, Google Transit API and Google Map API were used to build the velocity profile of a bus fleet operating on the NYC Metropolitan Transportation Authority (MTA) M1 route. The velocity profile varies depending on operation time, which results in a more realistic optimal result.
- The suboptimal design of a wireless charging electric bus system based on reinforcement learning was modeled to find the optimal values of battery capacity, pickup capacity, and the number of power-cable installations.
- A simulation of the proposed model was conducted for both static and dynamic traffic environments.

## 2. Modeling of Wireless Charging Electric Bus System

The system model for the wireless charging electric bus consists of two main elements, namely, the operating traffic environment, and the dynamics of the wireless charging electric bus. The overall model of the system is shown in Figure 2.

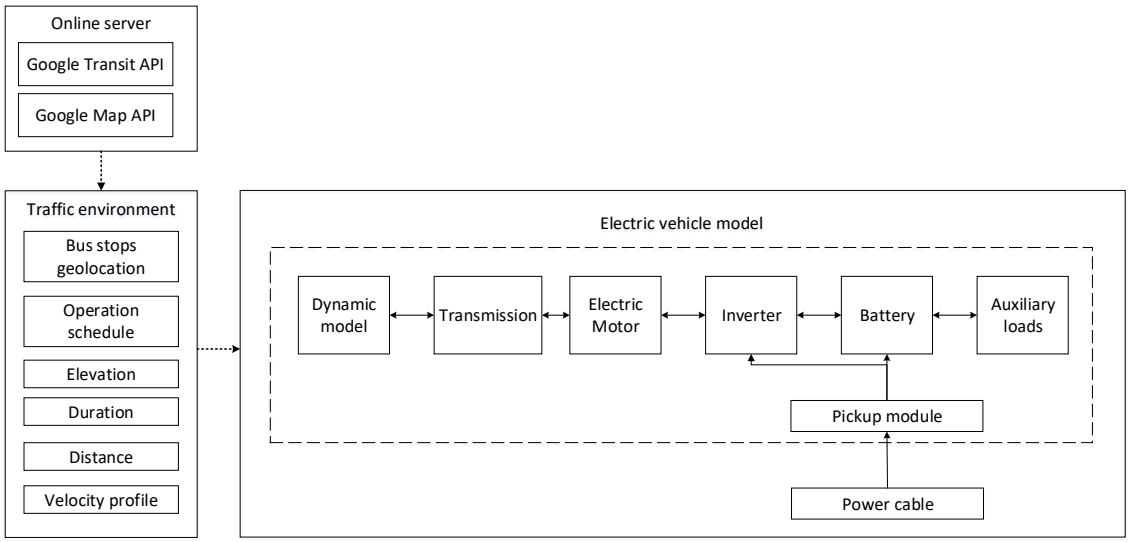


**Figure 2.** System configuration of wireless charging electric bus system.

### *2.1. Environment*

The wireless charging electric bus was simulated using the bus fleet of route M1 of the New York City MTA, which operates from Harlem to the East Village in Manhattan [21].

The NYC MTA M1 route was selected due to its distinct traffic environment depending on the time of day, as shown in Figure 3. The dynamic traffic environment is characterized by the diversity of the velocity profile of the route, in which the velocity of the electric bus varies depending on the time of operation. For example, mean acceleration and velocity during commuting hours are significantly lower due to heavy traffic. Low operational velocity means that the traffic environment is more conducive to the installation of power tracks, because low velocity increases charging time. When traffic is moving freely, operational velocity increases and it is beneficial to use an electric bus with a larger battery capacity compared with installing wireless charging infrastructure. To obtain the diverse characteristics of the velocity profile in this case, the Google Transit API was used to geolocate bus stations. Then, the distance and travel time between bus stations were found by using Google Map API. The following steps were used to build the diverse velocity profile:

- Geolocation information (latitude and longitude) for all 64 stops was found.
- Departure and arrival times were found for neighboring stations. Mean velocity was calculated using time differences and distances between all 64 stations.
- Velocity profile was constructed using mean velocity, deceleration, and acceleration data for the electric bus.

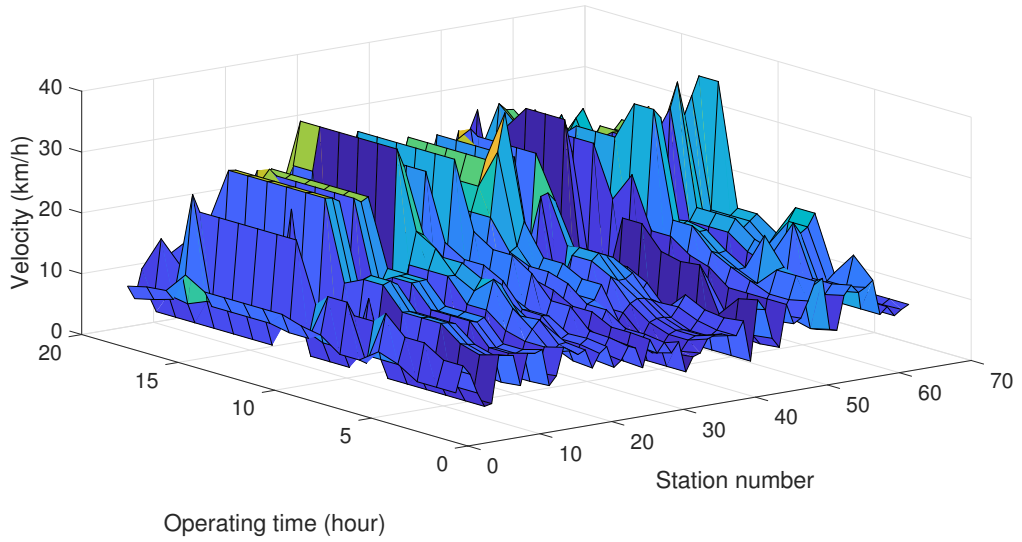


**Figure 3.** Velocity profile of NYC Metropolitan Transportation Authority (MTA) M1 line.

As a result, the velocity profile for the simulation was built, and examples of rush-hour (8:00) and nonrush-hour (15:00) velocity profiles are shown in Figure 4. We can observe that the velocity profile varies depending on the time of operation.

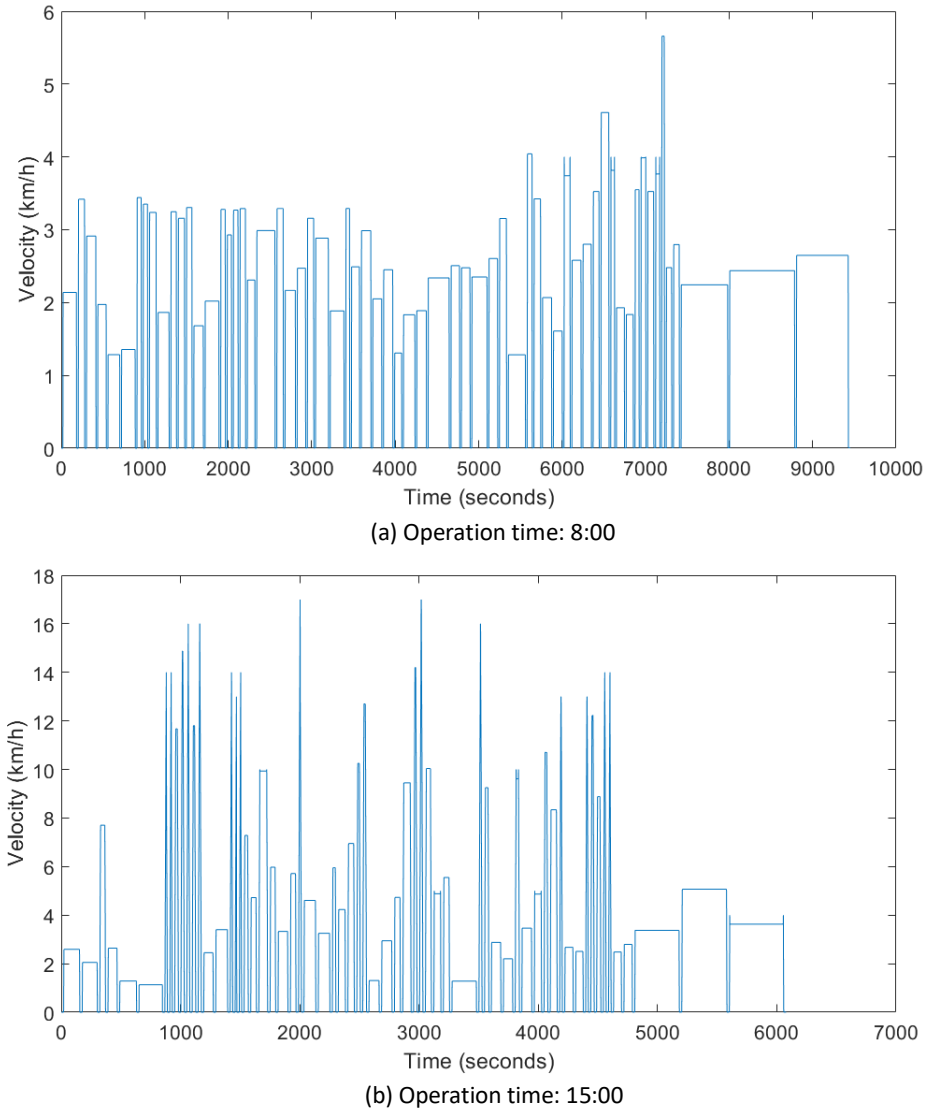


**Figure 4.** Velocity data based on operational timeline. (**a**) Operation time: 8:00; (**b**) Operation time: 15:00.

## *2.2. System Modeling*

### 2.2.1. Dynamic Characteristics of Wireless Charging Electric Bus

In this section, we consider the dynamic characteristics of the wireless charging electric bus that affect its performance and travel distance, as shown in Figure 5. First, tractive effort is the generated force when the electric bus is propelled forward, and energy is transmitted from the motor to the drive wheels [18]. Rolling resistance force is caused by friction between the vehicle wheel and the road. Rolling resistance force $F_{rr}$ is expressed as

$$F_{rr} = C_{rr} mg \tag{1}$$

Here, $C_{rr}$ is the rolling-friction coefficient, m is the mass of the vehicle, and g is the gravity force of 9.81 m/s$^2$. Aerodynamic drag force $F_{ad}$ is given by

$$F_{ad} = \frac{1}{2} \rho A C_d v^2 \tag{2}$$

where $\rho$ is air density, $A$ is the frontal area, $v$ is velocity, and $C_d$ is the drag coefficient. Hill climbing force $F_{hc}$ is proportional to vehicle weight and gravity force, which is expressed as

$$F_{hc} = mgsin(\theta) \tag{3}$$

where $sin(\varphi)$ is used to calculate the slope angle. Acceleration force $F_{la}$ provides the linear acceleration of the vehicle when the velocity of the vehicle varies with time. The tractive effort of the wireless charging electric bus can be described as

$$F_{te} = F_{rr} + F_{ad} + F_{hc} + F_{la} \tag{4}$$

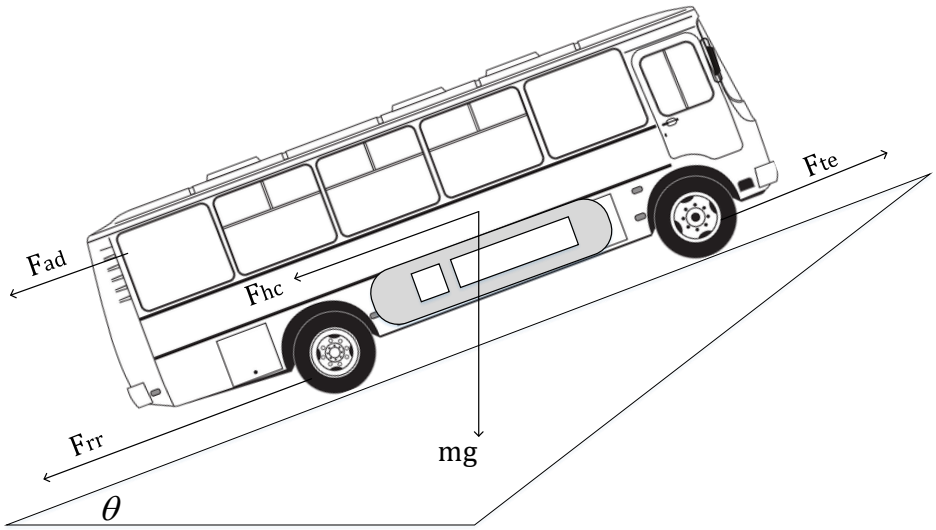


**Figure 5.** Dynamic characteristics of wireless charging electric bus.

2.2.2. Transmission

Transmission is modeled by considering the transmission power from the electric motor to the wheels, which governs the motion of the wireless charging electric bus under different driving conditions. The first step is to calculate traction torque $\tau_w$ of each wheel, which is expressed as [22]:

$$\tau_w = \frac{F_{te} r_w}{2} \tag{5}$$

where $r_w$ is the wheel radius. Depending on tractive power $P_{te}$, which is calculated by multiplying velocity to $F_{te}$, shaft torque $\tau_s$ is calculated as follows

$$\tau_s = \eta_{ef} \frac{2\tau_w}{G}, \quad P_{te} < 0 \tag{6}$$

$$\tau_s = \frac{2\tau_w}{\eta_{ef} G}, \quad P_{te} \geq 0 \tag{7}$$

here, $G$ is the gear ratio and $\eta_{ef}$ denotes the efficiency of the electric machine. Finally, the shaft power of electric machine $P_s$ is shown as

$$P_s = \tau_s G w_w \tag{8}$$

where $w_w$ is the angular velocity of the wheel.

2.2.3. Electric Motor

For the propulsion of the wireless charging electric bus, the electric motor is modeled under the assumption of a permanent-magnet synchronous machine (PMSM). The first step of modeling PMSM is to define voltage in a DQ-frame [23], as shown below:

$$\begin{pmatrix} V_d \\ V_q \end{pmatrix} = \begin{pmatrix} R_s + \rho L_q & w_r L_d \\ -w_r L_q & R_s + \rho L_d \end{pmatrix} + \begin{pmatrix} i_q \\ i_d \end{pmatrix} + \begin{pmatrix} w_r \lambda_f \\ \rho \lambda_f \end{pmatrix} \tag{9}$$

where $R_s$ is stator-phase resistance, $L_q$ denotes Q-axis inductance, $L_d$ denotes D-axis inductance, $\rho$ is the derivative operator, and $\lambda_f$ is the flux linkage. Based on the Park transform three-phase voltages, $V_a$, $V_b$, and $V_c$ can be acquired by the following equation:

$$\begin{pmatrix} V_a \\ V_b \\ V_c \end{pmatrix} = \begin{pmatrix} cos(\theta) & sin(\theta) & 1 \\ cos(\theta - 2\pi/3) & sin(\theta - 2\pi/3) & 1 \\ cos(\theta + 2\pi/3) & sin(\theta + 2\pi/3) & 1 \end{pmatrix} \begin{pmatrix} V_d \\ V_q \\ V_o \end{pmatrix} \tag{10}$$

Thus, motor input power $P_{mot,in}$ is derived as

$$P_{mot,in} = \frac{3}{2}(V_d i_d + V_q i_q) \tag{11}$$

Motor power is limited based on motors' max power $P_{mot,max}$, which is expressed as

$$P_{mot,max} = \frac{P_s}{\eta_{EM}} \tag{12}$$

where $\eta_{EM}$ is the efficiency of the electric motor.

2.2.4. Inverter

The inverter transmits power between the PMSM and the battery by turning the switches on and off. The switches of the inverter produce resistance that leads to a power loss. The average power losses of one switch can be described as follows [24]:

$$P_{q,inv} = (\frac{1}{8} + \frac{m_i}{3\pi}) R_{q,inv} i_p^2 + (\frac{1}{2\pi} + \frac{m_i}{8} cos(\phi)) V_{q,th,inv} i_p \tag{13}$$

$$P_{d,inv} = (\frac{1}{8} - \frac{m_i}{3\pi}) R_{q,inv} i_p^2 + (\frac{1}{2\pi} - \frac{m_i}{8} cos(\phi)) V_{q,th,inv} i_p \tag{14}$$

where $i_p$ is the phase current, $\phi$ is the power factor angle, $R_{q,inv}$ is the inverter switch resistance, and $V_{q,th,inv}$ denotes the threshold voltage of the inverter switch. Assuming the switch and voltage losses are equal, the above equation can be simplified to give

$$P_{inv,loss} = \frac{3}{2} R_{inv} i_p^2 + \frac{6}{\pi} V_{th,inv} i_p \tag{15}$$

Thus, the sum of $P_{inv,loss}$ and $P_{mot,in}$ gives the final total power that battery is required to supply as

$$P_{bat,in} = P_{inv,loss} + P_{mot,in} \tag{16}$$

2.2.5. Battery

The purpose of the battery model is to predict the performance of the wireless charging electric bus in terms of its range, acceleration, speed, and other vehicle parameters. To evaluate EV performance, a modified version of the Shepherd model [25,26] was used to evaluate the State of Charge (SoC)

level, which is a unit for evaluating the remaining battery charge. The battery life-cycle model can be expressed as

$$Q = 30.33e^{(-31.5/8.31T)}RS_{bat} \tag{17}$$

where $R$ is the number of cycles, $T$ denotes temperature, and $S_{bat}$ is the initial capacity of the battery. In the battery model, a controlled voltage source with internal resistance was used to obtain battery voltage $V_{bat}$. The calculation of controlled voltage source $E$ is described as follows

$$E = E_0 - K\frac{Q}{Q - C} + A\exp^{-BC} \tag{18}$$

where $K$ denotes the polarization voltage, $E_0$ is the battery constant voltage, $Q$ denotes battery capacity, $A$ expresses the amplitude of the exponential zone, and $B$ denotes the inverse of the exponential zone time, which is added to reflect the nonlinear characteristics of the lithium-ion battery. The specification of the lithium-ion battery is shown in Table 1, and the variation in battery voltage based on the actual battery charge is shown in Figure 6.

**Table 1.** Lithium-ion battery with 3.6 V and 6.2 Ah specification.

| Parameters | Value |
|---|---|
| Polarization voltage (V), $K$ | 0.00876 |
| Battery constant voltage (V), $E_0$ | 3.7348 |
| Battery capacity (Ah), $Q_init$ | 6.2 |
| Exponential zone amplitude, $A$ | 0.468 |
| Exponential zone time constant inverse $Ah^{-1}$, $B$ | 3.5294 |
| Temperature Celsius, $T$ | 25 |

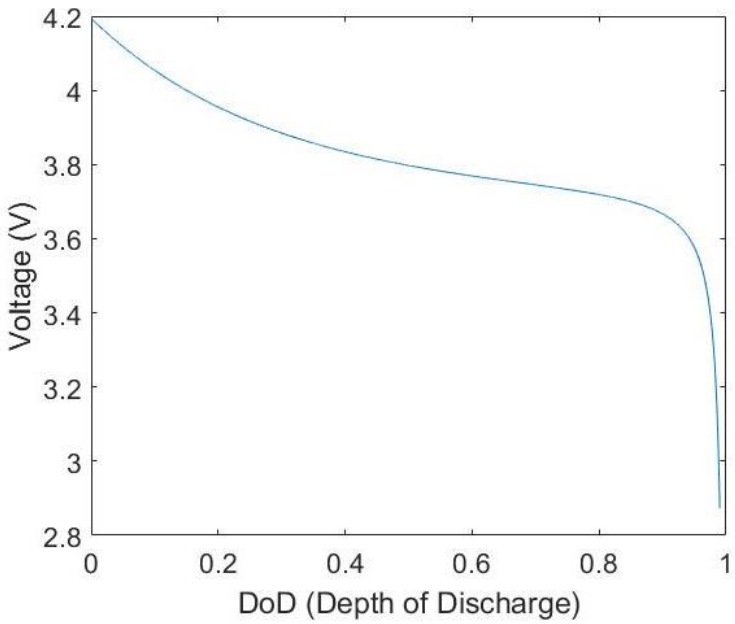


**Figure 6.** Battery-voltage variance based on depth of discharge (DoD).

After acquiring controlled voltage source $E$, battery voltage $V_{bat}$ is calculated according to the following equation:

$$V_{bat} = E - iR \tag{19}$$

Here, $i$ is the battery current and $R$ is the internal resistance. Battery power $P_{bat}$ is expressed by multiplying the terminal voltage by the output current as follows:

$$P_{bat} = iV_{bat} = i(E - iR) = iE - i^2R \tag{20}$$

Then, using Equation (20), battery output current $i$ is described by

$$i = \frac{E - \sqrt{E^2 - 4RP_{bat}}}{2R} \tag{21}$$

When the vehicle comes to a stop or its speed decreases, the direction of the output current changes to dissipate power into the vehicle's battery, which is expressed as

$$i = \frac{-E + \sqrt{E^2 - 4RP_{bat}}}{2R} \tag{22}$$

The next step is to update the battery capacity, which changes according to its output current, as shown below:

$$CD_{n+1} = CD_n + Q \tag{23}$$

Here, $CD_n$ denotes the disposed charge at the $n$th step of the simulation, and $Q$ represents initial battery capacity. The final step is to update the $SoC$ of the battery. At the n-th step, the $SoC$ level of the simulation, $soc_n$ can therefore be represented as

$$soc_n = 1 - \frac{CD_n}{Q} \tag{24}$$

$soc_n$ is then used to measure the battery charge level during simulation.

#### 2.2.6. Wireless Charging Module

The wireless charging module was designed to supply power to the wireless charging electric bus. The Series–Series (SS) compensation-network model is used on both the primary and secondary sides of the wireless charging system. The high-frequency current on the primary side generates an alternating magnetic field to induce voltage on the secondary side [27,28]. In order to precisely calculate the transferred power from the power cable to the pickup module, maximum power efficiency is calculated by the equation

$$\eta_{max} = \frac{k^2 Q_1 Q_2}{(1 + \sqrt{(1 + k^2 Q_1 Q_2)})^2} \tag{25}$$

where $k$ is the coupling coefficient, and $Q_1$ and $Q_2$ are the quality factors for the primary and secondary sides. Using $\eta_{max}$, the transferred power from the power cable to the pickup module is determined as:

$$P_{pm} = \eta_{max} PC_{loc} \tag{26}$$

Then, $P_pm$ is redirected in to charge the battery as shown below:

$$P_{bat} = P_{bat} + P_{pm} \tag{27}$$

## 3. Suboptimal Design of Wireless Charging Electric Bus System Based on Reinforcement Learning

In this section, we introduce an optimization algorithm based on reinforcement learning to find the optimal battery capacity, pickup capacity, and number of power-cable installations. The proposed algorithm is built using an MDP, which consists of a set of finite states $s \in S$, a set of possible actions $a \in A$ in each state, a real-value reward function $r \in R$, and a state-transition model $P(s, a')$

[29,30]. The model used in this paper can be verified as a Markov model because state transitions are independent of any previous states or actions. The layout of the proposed optimization algorithm is shown in Figure 7.

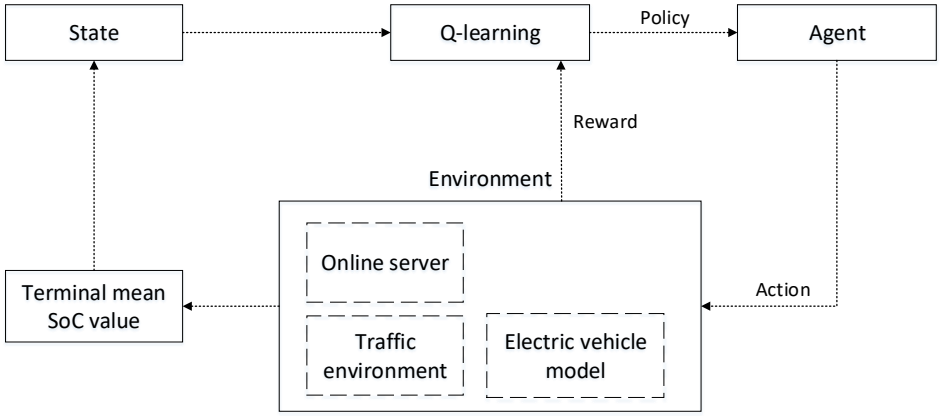


**Figure 7.** Architecture of proposed reinforcement-learning algorithm.

### *3.1. Action-State Value Update*

Q-learning was adopted to maximize the action value as the state changes from the current to the next state according to the action taken by the agent. As a result, for a given action, the Q value is updated based on immediate and future rewards obtained from the environment. An immediate reward is the reward for any recent change in the state brought about by the action of the agent, and a future reward is the reward associated with the future environment resulting from the action. Ultimately, the agent's ultimate goal is to update the value of Q to obtain the maximum reward as shown below:

$$Q_{t+1}(s_t, a_t) \leftarrow Q_t(s_t, a_t) + \alpha\{R(s_t, a_t) + \gamma max_{a \in A}(Q_t(s_{t+1}, a) - Q(s_t, a_t))\} \quad (28)$$

where $s$ is the state, $a$ is the action, and $r$ denotes the reward. $\gamma$ is a discount factor between 0 and 1; a value closer to 1 emphasizes the importance of compensation for the future. In this paper, $\gamma$ was set to 0.5 in order to respond to a diverse traffic environment. $\alpha$ is a learning rate with a value between 0 and 1, which determines the learning rate of the Q value. For example, if $\alpha = 0$, agent learning is disabled. If $\alpha = 1$, the agent learns using the most recent information. In this case, $\alpha$ was set to 1 because the agent must learn from the previous Q value. The Q value is thus updated in the multidimensional Q-table shown in Figure 8.

The first dimension denotes the state space. Then, the second, third, and fourth dimensions denote the action space, in which the change of variables (battery capacity, pickup capacity, power-cable installation number) are defined by action. The reasons for building multiple layers in the Q-table are as follows:

- During Q-learning, any increase or decrease in each variable can easily be checked.
- Each dimension links to each action: the change of battery capacity, pickup capacity, and power-cable installation number.
- The agent only needs to search the actions around the current state, not the whole Q table.
- After random sampling, exploitation converges much faster because each $Q$-value has its own unique domain.

Algorithm 1 represents the pseudocode of the proposed algorithm based on reinforcement learning. First, all entries in the multidimensional Q-table are initialized to zero. According to policy $\pi$, actions are selected for either exploitation or exploration. For the termination of the algorithm, $\delta$ was set to a small constant value. In the case of exploration, actions are randomly selected, and the

subsequent $r$ and $Q_t(s_t, a_t)$ are updated. In exploitation, actions that maximize future value $Q_{t+1}(s_t, a_t)$ are identified. For a given $a$, $r$, the $Q_t(s_t, a_t)$ are updated. Finally, if the difference between future $Q_{t+1}(s_t, a_t)$ and current $Q_t(s_t, a_t)$ is less than $\delta$, the process terminates.

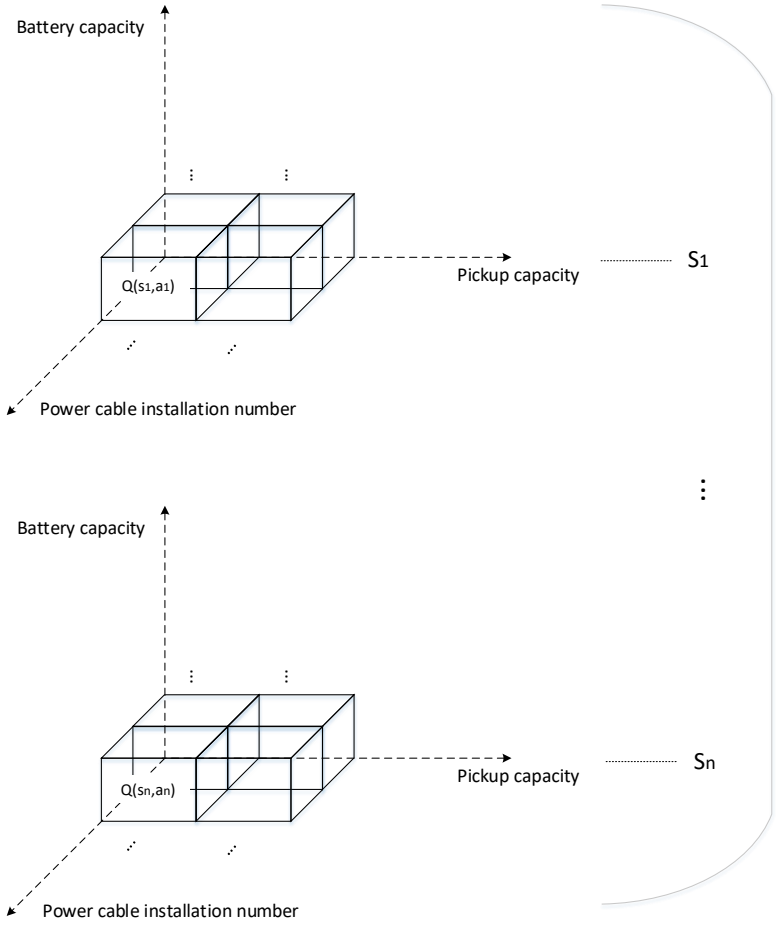


**Figure 8.** Multilayer Q-table for proposed optimization algorithm.

**Algorithm 1** Proposed optimization algorithm

Initialize $Q(s, a)$ for all $s \in S$ and $a \in A$
Initialize $\pi$ to be $\epsilon - greedy$ with respect to $Q(s, a)$
small $\delta > 0$
**if** $abs(Q_{t+1}(s_t, a_t) - Q(s_t, a_t)) > \delta$ **then**
  **for** $t$=0,1,2...$T$ **do**
    **if** *rand(1)* $< \epsilon$ **then**
      with probability $\epsilon$ select random action $a_t$
    **else**
      select action which maximizes the $Q$ value $a_t = argmax_{a' \in A} Q(s_t, a')$
    observe and store next reward $r_{t+1} = \frac{max\{C^t\} - C^t}{\eta} + \tau$
    calculate next state $S^{t+1}_{EV,soc} \in \{soc^1, soc^2, ...soc^t | S^t_{bat}, PC^t_{loc}\}$
    update next value
      $Q_{t+1}(s_t, a_t) \leftarrow Q_t(s_t, a_t) + \alpha\{R(s_t, a_t) + \gamma max_{a \in A}(Q_t(s_{t+1}, a) - Q(s_t, a_t))\}$
**else**
  terminate the process save final state $s$.

### *3.2. State*

The state is given by the average SoC of the bus fleet, which is calculated based on battery capacity, pickup capacity, and number of installed power cables. First, battery capacity $S_{bat}$ of a vehicle at step $t$ when it is accelerating can be written as:

$$S_{bat}^{t} \in \{B_{cap,1}, B_{cap,2} ... B_{cap,N_{bat}}\} \tag{29}$$

$$\frac{S_{bat}^{t} - P_{te}^{t} + \sum_{loc=1}^{N_{loc}} (t_{end}^{i} - t_{start}^{i}) CP^{t}}{B_{cap,N_{bat}}} \geq S_{bat,min} \tag{30}$$

where $t_{end}^{i} - t_{start}^{i}$ denotes the recharging time at the bus stop, which is randomly allocated at between 20 and 40 s, $CP$ denotes the pickup capacity, $N_{loc}$ represents the total number of power cables, and $B_{cap,N_{bat}}$ denotes the full battery capacity. For deceleration or downhill travel, regenerative breaking charges the battery, in which its maximum capacity is limited to $S_{bat,max}$, expressed as:

$$\frac{S_{bat}^{t} + P_{te}^{t} + \sum_{loc=1}^{N_{loc}} (t_{end}^{i} - t_{start}^{i}) CP^{t}}{B_{cap,N_{bat}}} \leq S_{bat,max} \tag{31}$$

If the bus stop is located in the area with a low operation speed, it has priority over power-cable installation. The installed power cables in the bus-stop locations are shown as

$$loc^{t} \in \{L_{1,pr}, L_{2,pr}, ... L_{i,pr}\} \tag{32}$$

where $L_i$ represents a location of a bus station, and $pr$ shows the priority of the according bus stations. The power cable is installed at the bus stop with designated pickup capacity, described as

$$PC_{loc}^{t} \in \{pc_{l}^{1}, pc_{l}^{2}, ... pc_{N}^{t} | CP^{t}\} \tag{33}$$

where $pc_{N}^{t}$ represents the location of a power cable with given pickup capacity $CP^{t}$. The electric bus follows velocity profile $V^{t}$ that changes according to time step $t$. From this velocity profile, SoC $EV_{soc,n}^{t}$ of the wireless charging electric bus is calculated for each time step $t$, expressed as

$$EV_{soc,n}^{t} = EV_{sys}(V^{t}, S_{bat}^{t}, PC_{loc}^{t}) \tag{34}$$

where $EV_{sys}$ is the wireless charging electric bus system model introduced earlier. The charging behavior of a wireless charging electric bus is based on the location of a power cable and pickup capacity, as shown below:

$$EV_{soc,loc} = EV_{loc}^{t} \bigcap PC_{loc}^{t} \tag{35}$$

where $EV_{soc,loc}$ is charged when there is an installed power cable at the station. When the wireless charging electric bus fleet ends all its operations, the average $soc$ is calculated for all electric buses as:

$$SoC_{avg} = \frac{\sum_{n=1}^{N} soc_{EV}^{n}}{n} \tag{36}$$

With a given battery capacity, pickup capacity, and number of power cables, the average $SoC_{avg}$ is then saved to the state after each episode, expressed as

$$S_{EV,soc} \in \{soc^{1}, soc^{2}, ... soc^{t} | S_{bat}^{t}, PC_{loc}^{t}\} \tag{37}$$

### 3.3. Action and Reward

After defining the state, action set $A$ is formulated by selecting the three main variables, $S_{bat}^t$, $PC_{loc}^t$, and $CP^t$. We define $a_t$ as the action taken at time step $t$, and the following actions are selected from possible action set $A$. All possible actions $a^t$ are defined as

$$a^t = \begin{cases} S_{bat}^t + \Delta_1 \\ PC_{loc}^t + \Delta_2 \\ CP^t + \Delta_3 \end{cases} = [\Delta_1, \Delta_2, \Delta_3] \tag{38}$$

where variables increase or decrease each other variable by a value $\Delta_n$, which is determined based on the size of the environment's state space. To increase accuracy, $\Delta_n$ can be reduced, and, to improve computational speed, $\Delta_n$ can be increased. The total cost of the wireless charging system can be derived for the reward function as shown below:

$$C^t = \{n_v(p_v + S_{bat}^t p_b) + (\sum_{PC_{loc}=1}^{N_{loc}} (t_{end}^{PC_{loc}} - t_{start}^{PC_{loc}}))p_l + N_{loc} CP p_i\} \tag{39}$$

where $n_v$ is the number of operating wireless charging electric buses, $p_v$ denotes the price of a wireless charging electric bus, $p_b$ is battery price, and $p_l$ represents the cost of the power cable. The first term of $C^t$ indicates the cost of the wireless charging electric bus system, including the battery and pickup module, while the second term represents the total cost of the power cables installed along the route. For total cost $C^t$, the reward is expressed as

$$R^t = \frac{max(C^t) - C^t}{\eta} + \tau \tag{40}$$

where $\tau$ is determined based on state $S_{EV,soc}$, and $\eta$ is the scaling factor. If $S_{EV,soc}$ is greater than a 0.2 SoC threshold level, $\tau$ is set to 1. Otherwise, $\tau$ is set to $-1$. The main aims of the proposed algorithm are to ensure a stable operation and to reduce the total investment cost. The inclusion of $\tau$ can increase the likelihood of selecting an action with a low cost, but which threatens the stable operation of the wireless charging electric bus fleet. The reward is therefore based on two factors, namely, total cost and the stable operation of the wireless charging electric bus.

## 4. Results and Discussion

### 4.1. Simulation Environment

In the following section, we describe how an optimized value is obtained using a reinforcement-learning algorithm. We therefore need to simulate the real environment to verify that the variables from the selected state are close to optimal. In order to assess the proposed algorithm, we prepared a scenario using the wireless charging electric bus fleet operating on the NYC MTA M1 route. The details of the simulation environment are summarized below:

- Simulation begins when $n_v$ dynamic charging electric buses move forward from their starting points.
- Each electric bus departs with a fully charged battery and stops at each station for 20 to 40 s.
- The number of passengers boarding the bus differs over the timeline, and this, in turn, affects the total weight of the bus. The number of passengers peaks during commuting time and gradually reduces.
- The velocity-profile changes and the data for each episode are directly obtained from Google Map API.
- The route length is fixed, and the journey ends when the wireless charging electric bus returns to its starting point.

- The bus receives power from a single source, namely, the power cables installed underground.

Figure 9 shows the overall simulation environment described above. The parameters of the simulation environment are shown in Table 2.

**Table 2.** Parameters of the simulation environment.

| Parameters | Cost ($) |
|---|---|
| Price of battery capacity ($/kWh), $p_b$ | 290 |
| Price of inverter capacity ($/kW), $p_i$ | 120 |
| Price of power cable segment ($/No.), $p_l$ | 5000 |
| Price of electric bus ($/No.), $p_v$ | 160,000 |
| Route length (km) | 20.2 |

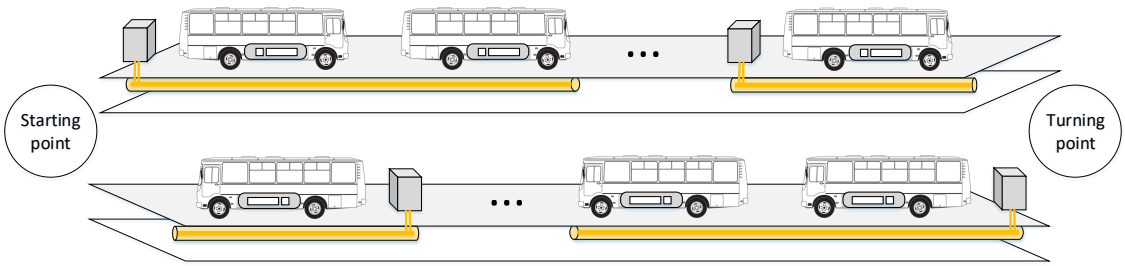


**Figure 9.** Layout of simulation environment.

### *4.2. MIP-Based Exact Algorithm*

The MIP-based exact algorithm is used as a benchmark to compare the outputs from proposed reinforcement-learning algorithm. The objective function of the wireless charging electric bus system is expressed as

$$min\{n_v(p_v + S^t_{bat}p_b) + (\sum_{PC_{loc}=1}^{N_{loc}}(t^{PC_{loc}}_{end} - t^{PC_{loc}}_{start}))p_l + N_{loc}CPp_i\} \tag{41}$$

The first term of Equation (41) indicates the cost of the wireless charging electric bus, including the battery and pickup module, while the second term represents the total cost of the installed power cables along the route. The cost associated with pickup capacity is shown in the third term.

Subject to:

$$S^t_{bat} \in \{B_{cap,1}, B_{cap,2}...B_{cap,N_{bat}}\},\ t \in \{1,2,...T\} \tag{42}$$

$$\frac{S^t_{bat}(0)}{B_{cap,N_{bat}}} = 1,\ t \in \{1,2,...T\} \tag{43}$$

$$CP^t \in \{C_{pic,1}, C_{pic,2}...C_{pic,n}\},\ t \in \{1,2,...T\} \tag{44}$$

$$PC^t_{loc} \in \{pc^1_l, pc^2_l, ...pc^t_{N_{loc}}|CP^t\},\ t \in \{1,2,...T\} \tag{45}$$

$$PC^t_{loc} \le N_{station},\ t \in \{1,2,...T\} \tag{46}$$

$$\frac{S^t_{bat} - P^t_{te} + \sum_{loc=1}^{N_{loc}}(t^i_{end} - t^i_{start})CP^t}{B_{cap,N_{bat}}} \ge S_{bat,min},\ t \in \{1,2,...T\} \tag{47}$$

$$\frac{S^t_{bat} + P^t_{te} + \sum_{loc=1}^{N_{loc}}(t^i_{end} - t^i_{start})CP^t}{B_{cap,N_{bat}}} \le S_{bat,max},\ t \in \{1,2,...T\} \tag{48}$$

$$n^t_v \ge n_{v,min},\ t \in \{1,2,...T\} \tag{49}$$

$$EV_{sys}(V^t, S^t_{bat}, PC^t_{loc}) \ge S_{bat,min},\ t \in \{1,2,...T\} \tag{50}$$

In Equation (42), the battery capacity of electric bus is defined. Then, all the electric buses have equivalent battery capacity $S_{bat}$ for the following time step $t$. Before the vehicle starts moving, battery capacity must be full, as shown in Equation (43). In Equation (44), the pickup capacity of a power cable is determined to charge the electric bus at the station. The power cable is installed at the bus stop with a designated pickup capacity as expressed in Equation (45). Moreover, the number of power cables is limited to the total number of bus stations $N_{station}$. The electric bus' battery capacity is affected by the tractive effort and the power cables installed at the bus station. When the electric bus is operating on a power cable, the battery is charged in proportion to time ($t_{end} - t_{start}$) at the station, where $t_{start}$ denotes the charge starting time, and $t_{end}$ expresses the charge end time on the power cable. If the electric bus is accelerating, the electric bus' battery is discharged. In this case, the minimum battery capacity is limited to $S_{bat,min}$, which is shown in Equation (47). For deceleration or downhill travel, regenerative breaking charges the battery, and the battery maximum capacity is limited to $S_{bat,max}$, which is expressed in Equation (48). Equation (49) shows that the number of vehicles must be greater than $n_{v,min}$, which is the minimum number of operating vehicles required to satisfy the operation schedule of the electric bus. Finally, Equation (50) shows that the remaining battery capacity of the electric bus must be greater than the minimum battery-capacity threshold.

### *4.3. Convergence of Proposed Optimization Algorithm*

Finding the optimal epsilon value is crucial because it determines the proportion of exploitation and exploration, which affects algorithm performance. Exploitation implies that the learner chooses the best action from the previous history, while exploration means that the learner chooses an arbitrary action to update the history. Thus, exploration increases the number of action-state values and the probability of finding a maximum reward. However, if exploration exceeds a certain value, the reward diverges and, eventually, $Q$-value falls into a local minimum. Therefore, it is important to balance exploitation and exploration using an epsilon value. To obtain the optimal epsilon value, the Q-table was initialized using random samples for up to 500 episodes. After 500 episodes, epsilon values between 0 and 1 were tested to assess how $\epsilon$-greedy affects the optimization process.

In Figure 10, the average reward is shown for epsilon values between 0 and 1. When the epsilon value is set to 0, the average reward value is found at a single value of 1.34. This is because the proposed algorithm starts to exploit the maximum $Q$-value without the need for any exploration. Therefore, the difference between the average rewards is minimal. As epsilon increases to 0.2, with the appropriate amount of exploration, the median value of the average reward reaches its highest point. Moreover, the number of outliers considerably increases due to the gap between a random action and an action selected based on enlarging the reward. After reaching a maximum average reward at an epsilon of 0.2, the average reward decreases until epsilon reaches 1. The average reward decreases because high exploration prevents the agent from choosing an action that maximizes the $Q$-value. An expansion in the whisker range can also be seen, because the average reward is spread over a large range. As a result, we conclude that an epsilon value of 0.2 maximizes the average reward in this case.

Figure 11 shows SoC variation for operating the wireless charging electric bus. When there is only exploitation, the SoC converges to 1. Although a stable operation is ensured, overinvestment in batteries and wireless charging infrastructure can be seen in Figure 12. To ensure a stable operation and to minimize total cost at the same time, SoC should be close to the SoC threshold of 0.2. When epsilon equals 0.2, the SoC range is smaller compared to the usage of other epsilon values. It can be seen that most SoC values are clustered between 0.2 and 0.4. This result shows that investment in battery capacity, pickup capacity, and power-cable installation is well-balanced. As a result, the minimum total cost can be found, as shown in Figure 12, for an epsilon of 0.2. When epsilon is increased to 0.4, the SoC values of the operating buses are lower than the threshold SoC level of 0.2, which makes it impossible to have a successful operation.

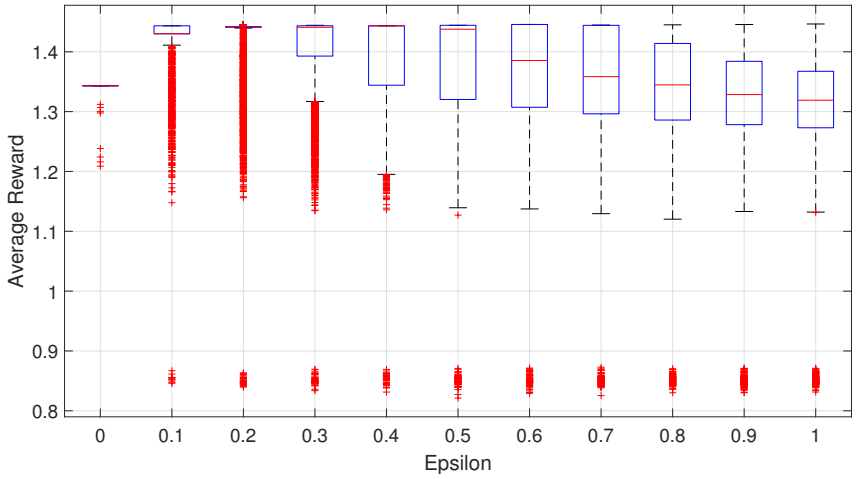


**Figure 10.** Average reward for different epsilon values.

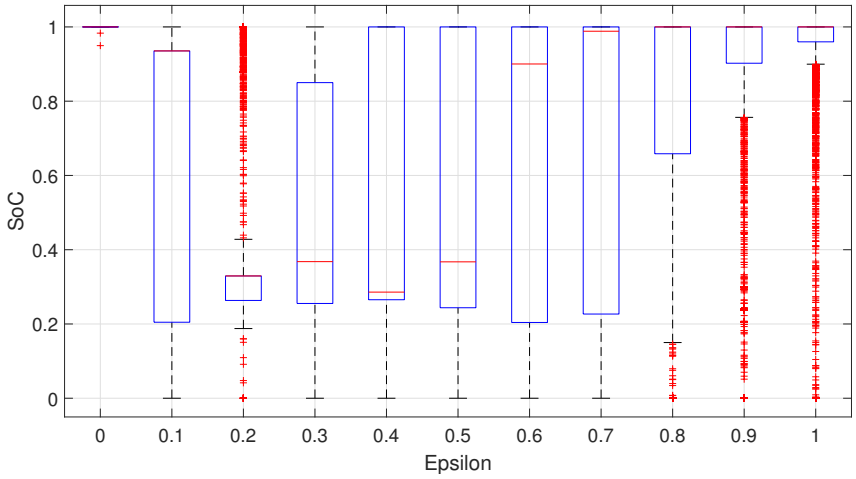


**Figure 11.** State of Charge (SoC) for different epsilon values.

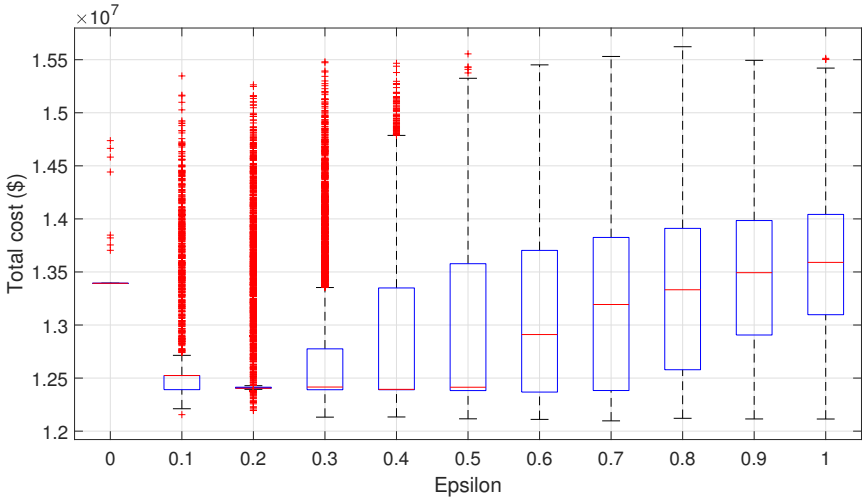


**Figure 12.** Total cost for different epsilon values.

### *4.4. Analysis in a Static Traffic Environment*

For a static traffic environment, a simulation was conducted with a constant velocity profile and the same number of passengers during operation. Figure 13 shows a drastic fall in total cost during the early stages for the proposed algorithm. Because a limited number of states is generated from

the selected action set, a state with low total cost was selected with a high percentage at the selection stage. In the latter stages of the proposed algorithm, it can be seen that fluctuation in total cost is reduced, and convergence begins after around 1000 episodes. This is because states with a high total cost are filtered out due to the low reward. As a result, states containing optimal values of battery capacity, pickup capacity, and number of installed power cables can be found after 1000 episodes. To verify whether the acquired total cost is the minimal total cost, the result of the proposed algorithm was compared with the total cost of the exact algorithm based on MIP. Figure 13 indicates that the minimal total-cost difference between the proposed algorithm and the MIP-based exact algorithm can be ignored.

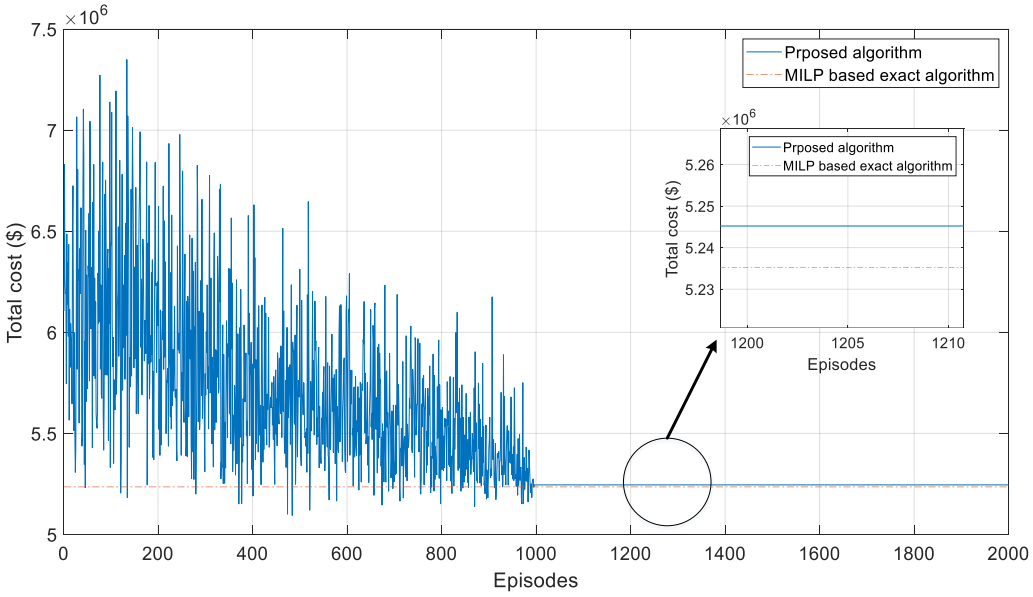


**Figure 13.** Cost comparison of exact and proposed algorithms in a static traffic environment.

### *4.5. Analysis in a Dynamic Traffic Environment*

In a dynamic traffic environment, some difficulty arises in estimating the power-consumption trajectory over the route due to uncertainty caused by changes in traffic and in the number of passengers on the bus. Moreover, it becomes hard to use the MIP-based exact algorithm because the optimal value changes after each episode as the traffic environment changes. It is, therefore, difficult to find the optimal variables using the exact algorithm for a continuous simulation period.

In Figure 14, we observe that the the MIP-based exact algorithm failed to converge as traffic environment changed after each episode. In contrast, the proposed algorithm is based on modelfree reinforcement learning, meaning that it only depends on feedback from the environment. Since the proposed algorithm was modeled to consider both current and future rewards, an action is selected to maximize the reward considering future changes in a traffic environment. Although a greater number of additional episodes is required to achieve convergence compared to a static traffic environment, the total cost finally reaches convergence after 2500 episodes.

Compared to a static traffic environment, total cost is increased by 39% because the number of operating wireless charging electric buses and pickup capacity are increased to ensure stable operation in dynamic traffic environments, as shown in Table 3. As the velocity profile changes hour by hour, the number of operating buses escalates to maintain the batch interval time. For instance, overall operating velocity decreases during rush hour, which results in longer operation time. To address this problem, additional buses are needed to maintain the batch interval. Variation in battery capacity is minimal because increasing battery capacity leads to an increase in recharging time. Therefore, additional buses would be required to fill the gap caused by longer recharging times. On the other hand, pickup capacity and power-cable installation number increased by a large margin compared

with a static traffic environment. This is due to the fact that, by increasing pickup capacity and power-cable installation, the number of operating buses can be minimized. Each bus can be equipped with a small battery because high pickup capacity can cause the battery to recharge in a relatively short time. Therefore, the bus can immediately be deployed for operation, which minimizes the number of additional buses required.

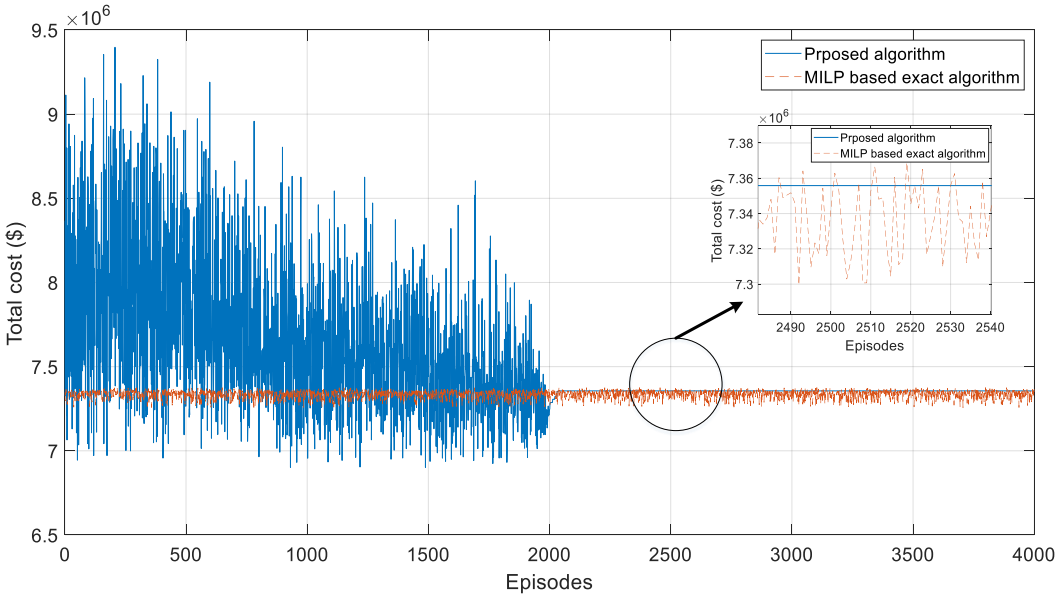


**Figure 14.** Cost comparison of exact and proposed algorithms in a dynamic traffic environment.

**Table 3.** Comparison between static and dynamic traffic environments.

| Variables | Static Traffic Environment | Dynamic Traffic Environment |
|---|---|---|
| Number of operating buses | 25 | 34 |
| Battery capacity | 24 kWh | 29 kWh |
| Pickup capacity | 77 kW | 138 kW |
| Number of installed power-cable segments | 5 | 12 |

## 5. Conclusions

Wireless charging is an innovative system that can be used to overcome the limitations of conventional electric buses. The advantage over conventional electric vehicles lies in its ability to recharge during motion. This feature is extremely attractive in comparison with conventional electric-vehicle systems. In the wireless charging electric bus system, the three components of pickup, battery, and number of power-cable installations all have considerable impact on the total cost of the system. In order to implement an efficient wireless charging electric bus system, we proposed a wireless charging electric bus system model and an optimization algorithm based on reinforcement learning to derive the optimal values for these main components. To make our analysis as realistic as possible, a dynamic traffic environment was modeled based on real traffic data of the NYC MTA M1 route, obtained using Google Transit API and Google Map API. The proposed model was built using a Markov decision process, composed of environment, state, action, reward, and policy. The same model was also used to verify the optimal numbers of power cables, pickup capacity, and battery capacity. Reinforcement learning was applied to solve the optimization problem of the wireless charging electric bus system in a diverse traffic environment. Numerical results showed that the proposed algorithm maximizes the average reward and minimizes the total cost by balancing the SoC close to the threshold. Furthermore, the convergence of the proposed algorithm was verified against the outcome using an exact solution based on MIP.

**Author Contributions:** H.L. proposed the main idea and performed the modeling, simulation, and data analysis, and wrote the original draft; D.J. provided the materials; D.-H.C. reviewed and edited the paper. All authors contributed via discussions and participated in writing the manuscript.

**Funding:** This work was supported by the Institute for Information and Communications Technology Promotion (IITP), grant-funded by the Korea Government (MSIT) (2017-0-00708, Development of a simultaneous wireless-information and power-transfer system using the same resource based on multiple antennas).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Suh, N.P.; Cho, D.H. Making the Move: From Internal Combustion Engines to Wireless Electric Vehicles. In *The On-line Electric Vehicle*; Springer: Cham, Switzerland, 2017; pp. 3–15.
2. Kan, T.; Nguyen, T.D.; White, J.C.; Malhan, R.K.; Mi, C.C. A new integration method for an electric vehicle wireless charging system using LCC compensation topology: Analysis and design. *IEEE Trans. Power Electron.* **2017**, *32*, 1638–1650. [CrossRef]
3. Bi, Z.; Kan, T.; Mi, C.C.; Zhang, Y.; Zhao, Z.; Keoleian, G.A. A review of wireless power transfer for electric vehicles: Prospects to enhance sustainable mobility. *Appl. Energy* **2016**, *179*, 413–425. [CrossRef]
4. Li, S.; Liu, Z.; Zhao, H.; Zhu, L.; Shuai, C.; Chen, Z. Wireless power transfer by electric field resonance and its application in dynamic charging. *IEEE Trans. Ind. Electron.* **2016**, *63*, 6602–6612. [CrossRef]
5. Tan, L.; Guo, J.; Huang, X.; Liu, H.; Yan, C.; Wang, W. Power Control Strategies of On-Road Charging for Electric Vehicles. *Energies* **2016**, *9*, 531. [CrossRef]
6. Kim, H.; Song, C.; Kim, D. H.; Jung, D. H.; Kim, I. M.; Kim, Y. I.; Kim, J. Coil design and measurements of automotive magnetic resonant wireless charging system for high-efficiency and low magnetic field leakage. *IEEE Trans. Microw. Theory Tech.* **2016**, *64*, 383–400. [CrossRef]
7. Suh, N.P., Cho, D. H. *The On-Line Electric Vehicle: Wireless Electric Ground Transportation Systems*; Springer: Cham, Switzerland, 2017; pp. 20–24
8. Mi, C.C.; Buja, G.; Choi, S.Y.; Rim, C.T. Modern advances in wireless power transfer systems for roadway powered electric vehicles. *IEEE Trans. Ind. Electron.* **2016**, *63*, 6533–6545. [CrossRef]
9. Manshadi, S.D.; Khodayar, M.E.; Abdelghany, K.; Üster, H. Wireless charging of electric vehicles in electricity and transportation networks. *IEEE Trans. Smart Grid* **2018**, *9*, 4503–4512. [CrossRef]
10. Bi, Z.; Keoleian, G.A.; Ersal, T. Wireless charger deployment for an electric bus network: A multi-objective life cycle optimization. *Appl. Energy* **2018**, *225*, 1090–1101. [CrossRef]
11. Jang, Y.J.; Jeong, S.; Lee, M.S. Initial energy logistics cost analysis for stationary, quasi-dynamic, and dynamic wireless charging public transportation systems. *Energies* **2016**, *9*, 483. [CrossRef]
12. Chokkalingam, B.; Padmanaban, S.; Siano, P.; Krishnamoorthy, R.; Selvaraj, R. Real-time forecasting of EV charging station scheduling for smart energy systems. *Energies* **2017**, *10*, 377. [CrossRef]
13. Luo, Y.; Zhu, T.; Wan, S.; Zhang, S.; Li, K. Optimal charging scheduling for large-scale EV (electric vehicle) deployment based on the interaction of the smart-grid and intelligent-transport systems. *Energy* **2016**, *97*, 359–368. [CrossRef]
14. Javaid, N.; Javaid, S.; Abdul, W.; Ahmed, I.; Almogren, A.; Alamri, A.; Niaz, I.A. A hybrid genetic wind driven heuristic optimization algorithm for demand side management in smart grid. *Energies* **2017**, *10*, 319. [CrossRef]
15. Deilami, S. Online coordination of plug-in electric vehicles considering grid congestion and smart grid power quality. *Energies* **2018**, *11*, 2187. [CrossRef]
16. Khan, S.U.; Mehmood, K.K.; Haider, Z.M.; Bukhari, S.B.A.; Lee, S.J.; Rafique, M.K.; Kim, C.H. Energy Management Scheme for an EV Smart Charger V2G/G2V Application with an EV Power Allocation Technique and Voltage Regulation. *Appl. Sci.* **2018**, *8*, 648. [CrossRef]
17. Alhazmi, Y.A.; Mostafa, H.A.; Salama, M.M. Optimal allocation for electric vehicle charging stations using Trip Success Ratio. *Int. J. Electr. Power Energy Syst.* **2017**, *91*, 101–116. [CrossRef]
18. Doan, V.D.; Fujimoto, H.; Koseki, T.; Yasuda, T.; Kishi, H.; Fujita, T. Allocation of Wireless Power Transfer System From Viewpoint of Optimal Control Problem for Autonomous Driving Electric Vehicles. *IEEE Trans. Intell. Transp. Syst.* **2017**, *19*, 3255–3270. [CrossRef]

19. Chopra, S.; Bauer, P. Driving range extension of EV with on-road contactless power transfer—A case study. *IEEE Trans. Ind. Electron.* **2013**, *60*, 329–338. [CrossRef]
20. Van Vreckem, B.; Borodin, D.; De Bruyn, W.; Nowé, A. A Reinforcement Learning Approach to Solving Hybrid Flexible Flowline Scheduling Problems. In Proceedings of the 6th Multidisciplinary International Conference on Scheduling: Theory and Applications (MISTA), Gent, Belgium, 27–29 August 2013.
21. New York City MTA. Available online: http://www.mta.info/ (accessed on 5 January 2019).
22. Mesbahi, T.; Khenfri, F.; Rizoug, N.; Chaaban, K.; Bartholomeues, P.; Le Moigne, P. Dynamical modeling of Li-ion batteries for electric vehicle applications based on hybrid Particle Swarm–Nelder–Mead (PSO–NM) optimization algorithm. *Electr. Power Syst. Res.* **2016**, *131*, 195–204. [CrossRef]
23. Williamson, S.S.; Emadi, A.; Rajashekara, K. Comprehensive efficiency modeling of electric traction motor drives for hybrid electric vehicle propulsion applications. *IEEE Trans. Veh. Technol.* **2007**, *56*, 1561–1572. [CrossRef]
24. Ahn, K.; Bayrak, A.E.; Papalambros, P.Y. Electric vehicle design optimization: Integration of a high-fidelity interior-permanent-magnet motor model. *IEEE Trans. Veh. Technol.* **2015**,*64*, 3870–3877. [CrossRef]
25. Gong, X.; Xiong, R.; Mi, C.C. A data-driven bias-correction-method-based lithium-ion battery modeling approach for electric vehicle applications. *IEEE Trans. Ind. Appl.* **2016**, *52*, 1759–1765.
26. Tremblay, O.; Dessaint, L.A.; Dekkiche, A.I. A generic battery model for the dynamic simulation of hybrid electric vehicles. In Proceedings of the IEEE Vehicle Power and Propulsion Conference, Arlington, TX, USA, 9–12 September 2007.
27. Zhang, W.; White, J.C.; Abraham, A.M.; Mi, C.C. Loosely coupled transformer structure and interoperability study for EV wireless charging systems. *IEEE Trans. Power Electron.* **2015**, *30*, 6356–6367. [CrossRef]
28. Li, S.; Mi, C.C. Wireless power transfer for electric vehicle applications. *IEEE J. Emerg. Sel. Top. Power Electron.* **2015**, *3*, 4–17.
29. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Petersen, S. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529. [CrossRef]
30. Wei, Q.; Lewis, F.L.; Sun, Q.; Yan, P.; Song, R. Discrete-time deterministic Q-learning: A novel convergence analysis. *IEEE Trans. Cybern.* **2017**, *47*, 1224–1237. [CrossRef]